**Introduction**

Proper nouns are a challenge for the apprehension of texts by machines. In information science and computational linguistics, out-of-vocabulary tokens (mostly common nouns absent from the dictionary, proper nouns, and foreign words) are delicate to deal with, especially in the case of historical corpora, since they require adjustments of indexing methods and fine-grained linguistic tools in order to be made queryable or visible. Nevertheless, one can assume that irregular, seldom seen words bear a particular significance and structure the texts in a peculiar fashion. Concerning toponyms, it is known from disciplines such as linguistics or anthropology that their relevance exceeds the frame of deictic and indexical functions, as they enfold more than a mere reference in space. His study on Bosavi language leads Feld (1996) to the idea that “place significance neither starts nor ends with the linguistic referentiality of placenames”, and that “there is considerable variation in how names hold and unleash significance”.

In the western tradition, a current of reflexion dating back to the 1960s has provided the theoretical foundations of the “spatial turn”, whose epitome is the concept of space as emergent rather than existing a priori, and composed of relations rather than structures (Warf 2009), as “our epoch is one in which space takes for us the form of relations among sites” (Foucault [1967] 1984). As a consequence, both definition and importance of space have been re-evaluated throughout the humanities. The emergence of currents named “GeoHumanities” (Dear et al. 2011) or “Spatial Humanities” (Bodenhammer et al. 2010), has prompted for a transfer of research objects between disciplines as well as an enforcement of the spatial turn in practice through specific methods of analysis. Their common denominator consists in opening up new spaces and experimenting in a transdisciplinary perspective in a field which has been evolving at an exponential pace since the last decade (Caquard & Cartwright 2014).

Concerning the maps themselves, some of the most compelling examples reflect a relativity in their construction and uses of maps: post-representational cartography (Rossetto 2014), where there is neither a ground truth, nor a cartographic truth, and where "the map is not objectively ‘above’ or ‘beyond’ that which is represented" (Crampton 2001). Although the maps seem immediately interpretable, they are not an objective result but a construct, the result of a filtering, "a connection made visible" (Moretti 1999). Cartography is said to go beyond the realization of static maps, to the description of emergent structures. As such, there can be no single or best map.

---

1 “The parallel disciplinary structure of German-language Literaturwissenschaft and linguistics is not merely anecdotal; rather, it is an index of the necessary interchange of methodologies and content which holds beneficial possibilities for both fields.” (Domínguez 2011)
Fulltext geocoding, also known as geoparsing or geographic information retrieval, describes the process of identifying place mentions in text and linking them to unambiguous spatial references; it typically involves three components: gazetteers, toponym detection and disambiguation. Methodological developments in this research area are tightly linked to progresses in mapping systems, mostly thanks to a technology-driven evolution (Juvan 2015), as Geographic Information Systems (GIS) and corresponding series of tools mostly come from other disciplines. An underlying assumption resides in the belief that understanding language and literature is not accomplished by studying individual texts, but by aggregating and analyzing massive amounts of data: because it is impossible for individuals to catch sight of everything in a large corpus, advocates of distant reading employ computational techniques to mine the texts for significant patterns and then use statistical analysis to make statements about those patterns (Wulfman 2014).

The article is a developed version of a conference presentation at the Digital Humanities conference (Barbaresi 2017); it is structured as follows: first, the digital collection is introduced along with methodological considerations, then the processes used to extract and visualize place names are presented, finally three different realizations are shown and commented: a map of toponyms and two different views of toponymic cooccurrences.

Die Fackel (1899-1936)

Karl Kraus (1874-1936) founded his own journal, Die Fackel ("The Torch"), in 1899 and published it until his death. This complex and unique work resists summary description in its whole and in detail. Die Fackel includes few illustrations; it first resembled newspapers and increasingly became a magazine that was privileged in its editorial independence. Although it originally featured a few articles by intellectual figures, after December 1911 (Nr. 338) Kraus became its nearly exclusive author. Its field of interest embraced all major debates of its time, such as politics, culture, journalism, war, and overall the spiritual impoverishment of the intellectual society. Die Fackel denounced social masks and hypocrisy, most notably in politics and among journalists, and contained a series of attacks on corruption (in particular of the Habsburg Monarchy), and nationalism (notably the pan-German movement).

One of the hallmarks of Kraus' words is that they are not always to be taken to the letter. His extensive work on words not only supports criticism and satire by exploiting the full extent of different meanings but also by letting the readers reflect on the impact of a series of discourses, influences and powers (Kouno 2015). For example in the first issue, Kraus is critical of journalists who exchanged their free will or free spirit (Freisinn) against a flair for free theater coupons (Freikartensinn). These plays on words are often not directly translatable, an easier accessible example of modification on word-level is his characterization of the first world war as a "chlorious offensive" (chlorreiche Offensive, Nr. 443) which conveys altogether a feeling for the vanity of the war, the major role of chemical weaponry, and a certain linguistic accent.

Beyond his puns, Kraus' prose is known for its elaborate structures, complex syntax and the numerous allusions and citations, whose density increases in the journal's later issues. His typical and idiosyncratic style comes to expression through a great variety of different forms, such as
essays, notes, commentaries, aphorisms, poems, and drama. His most characteristic satirical-literary technique is his montage and detournement of quotations, as he wittingly comments upon the quotes he finds in the press, in the literary works, and in the political speeches of his time.

**On mapping historical text collections in general and Die Fackel in particular**

In this paper, I present visualizations designed to analyze historical text collections in a spatial perspective. I wish to show that using place names and collocations as entry points can provide an additional, synthetic access to a digital edition of the work (AAC-Fackel) which is already available online. Although *Die Fackel* is particularly difficult to apprehend, in the digital edition the texts have been digitized, manually corrected as well as manually annotated with respect to the names of persons and institutions, so that most proper nouns which are not place names can be excluded from the study. I set out on a distant reading experiment leading to maps meant to uncover patterns and specificities which are not easily retraceable during close reading, in the sense that "mapping and visualisation concepts at the intersection of literary theory, cartography, and database technology could open up new dimensions of research." (Piatti et al. 2011). The methodological implications of this interdisciplinary work are discussed and exemplified by "distant cartographic reading" on corpus level:

"Thematic maps of spaces in literature visualize patterns that may tell the reader something important about the social content and impacts of literary forms; without these presumably 'non-theoretical' maps, such patterns would remain unnoticed and could not yield new theories." (Juvan 2015)

I focus on the concept of visualization, that is on the processes and not on the products (Crampton 2001), and present them together with a critical apparatus, by giving a theoretical perspective on what is being shown and seen. In fact, digital methods in humanities ought to be criticized, especially since the cartographic enterprise bears both a thrill and a risk: “adding more to the world through abstraction”, and “adding to the riskiness of cartographic politics by proliferating yet more renders of the world” (Gerlach 2014).

The plurinational nature of the Austro-Hungarian empire and the significance of Vienna as a cultural center at that time call for european-wide visualizations. In fact, *Die Fackel* is a European phenomenon with a decisively international extent, both with a thematic and a geographical focus. This work carries heterogeneity at its core and contains a considerable variety of toponyms (Biber 2001) which are highly significant because of the multinational nature of the Austro-Hungarian empire and the later formation of a territorially diminished state. Regional and national peculiarities are explicitly referenced, through the names or through cultural or linguistic characteristics such as word choices, diction and accents. There are for instance a significant number of judiciary chronicles that were discussed in German-speaking newspapers at that time and consequently dissected by Kraus. There are close ties to events happening in Bohemia, the beginning of

---

2 [http://aac.ac.at/fackel Free online access to 37 volumes, 415 issues, 922 numbers; more than 22.500 pages and 6 million wordforms.](http://aac.ac.at/fackel)
publication is co-existent with a prominent murder case in Czechia and Czech nationals are very often mentioned throughout the collection (Krolop 1994).

Place names can refer to people or events, they can also convey a particular function or message, like in the (very productive) construction of swear words, where three prominent types are found (Welzig 2008): substantive constructions with prepositional attributes: "Cherusci from Krems", "Teutons from Teplitz"; complex composita involving proper names, which may include modified components and are not analyzed here; and finally exemplary description of societal types and typology of discourses which are bound to nationality, socio-economic origins, confession, language, etc. Some toponyms are more easily extractable than others, the method described below does not apply to complex cases such as traditional denominations for parts of Vienna, which may be tricky to disambiguate even during close reading.

Mapping digital collections requires to describe the implications of making something visible, knowing that new visualizations are not always scientifically relevant or productive. Reading Die Fackel under the perspective of toponyms cannot end in maps which satisfy both literary and technical constraints. Maps are relative and subordinated to various purposes, the steps in making a map – selection, omission, simplification, classification, the creation of hierarchies, and 'symbolization' – are all inherently rhetorical (Harley 1989). As long as this background is not eclipsed by the apparent perfection of numbers, I believe it is not only possible but also beneficial to find new accesses to historical text collections. The complexity of interpreting Die Fackel has been compared to a minefield, as mistakes are practically unavoidable, but at the time time it is precisely the ambitious apprehension of this difficult terrain that makes it possible to confront the texts in a philological way, in short to "light up the minefield" (Krolop 1994).

**Extraction of toponyms**

Among the tendencies in geographic information retrieval and geocoding (Melo & Martins 2017), the extraction and normalization of named places, itineraries, or qualitative spatial relations, as well as the extraction of locative expressions are particularly relevant to study text collections. From the point of view of computational linguistics, toponyms are often out-of-vocabulary tokens. On lexical level, they are a potential error source for natural language processing tools, while on phrasal level, they are supposed to be identified by part-of-speech taggers as named entities (NE tag), or eventually by more fine-grained named entity recognition tools as place names. The processing chains usually stop at this point, they do not provide visualizations in the geographical sense, even if the toponyms can be linked to meta-information such as type and georeference.

In the field of information extraction, named entity recognition is a set of text mining techniques designed to discover named entities, connections and the types of relations between them. The particular task of finding place names in texts is commonly named place names extraction or toponym resolution. It involves the detection of words and phrases that may potentially be proper nouns as well as a second operation classifying them as geographic references (Nouvel et al. 2015). A further step, geocoding, resides in disambiguating and adding geographical coordinates to a place name (Leetaru 2012). Toponym resolution often relies on named-entity recognition and artificial
intelligence (Leidner & Lieberman 2011). However, knowledge-based methods using fine-grained data, for example from Wikipedia (Hu et al. 2014) or Wikidata (Vrandecic & Krötzsch 2014), have already been used with encouraging results. Combined approaches also lead to more complete or finer knowledge bases which are paramount in order to obtain fine-grained results.

Indeed, interdisciplinary work is paramount in order to produce cartographic visualizations of historical texts, as the necessity to adapt to different contexts (Alex et al. 2015) and to complement existing resources with a precise historical gazetteer (Borin et al. 2014) has been highlighted. However, corpus linguistics and geographical information systems have traditionally had very little to do with each other, although both approaches can benefit from each other (Gregory et al. 2011). The present endeavor grounds on a specially curated gazetteer: during the 20th century there have been significant political changes in Central Europe that have severely affected toponyms, so that geographical databases lack coverage and detail. Consequently, the database developed at the Austrian Academy of Sciences (Academy Corpora) in cooperation with the Berlin-Brandenburg Academy of Sciences (Language Center) focuses on Europe and follows from a combination of approaches. Gazetteers are curated in a semi-supervised way to account for historical differences. Manual compilation and screening is sometimes necessary, because of the different names in the main languages of the area or because of changes in recent history, gathering existing information on the Web and completing it in the light of historical German and Austrian maps. Additionally, concurrent out-of-vocabulary tokens such as frequent proper and literary names are listed in order to exclude them from the study. Current geographical information is used as a fallback: Wikipedia's API3 is used to navigate in categories and to retrieve coordinates, which are completed by hand for states and regions; current information is also compiled from the Geonames database4: data for European countries are retrieved and preprocessed (variants and place types).

The extraction is performed by a sliding window which captures single tokens as well as multi-word expressions: the tokenized files of works to be analyzed are filtered and matched with the database by finite-state automats. The recognition is refined through a series of contextual adaptations (Barbaresi 2017). As it is often necessary to assign the right coordinates to a toponym among several possible ones, disambiguation is a critical component (Leetaru 2012). It has been shown that an acceptable precision can be reached by applying a series of heuristics based on existing meta-information or various data generated on-the-fly (Pouliquen et al. 2006). The use of a specially designed register as described above allows for bypassing the disambiguation for a hand-picked list of places, which accounts for the specifics of the texts and the time. In the other cases, two different methods (Buscaldi 2011) are used so far: map-based (geographically relevant contextual information) and knowledge-based (supplied external meta-information). The information taken into account consists in type and importance of the entries (as known from data extracted from Geonames or Wikipedia such as population counts) as well as immediate context (e.g. the expected range and the last countries and locations seen), which are controlled by predefined parameters, most notably customized distance calculations (here the distance to Vienna), filter level or size of the search radius.

3 https://www.wikidata.org
4 http://www.geonames.org
Geographical and topical overview

In order to provide a typological overview of the digital collection, a cascade of filters is used: current and historical states (e.g. Austria-Hungary); regions, important subparts of states, and regional landscapes (e.g. Swabia); populated places; and geographical features (e.g. seas or valleys). The results are projected on a map of Europe which is static in nature, it features combined phenomena in a single image and provides an overview of toponyms in the collection. The use of a cartographical software (TileMill) allows for the manipulation of mapping procedures, experimentation and fine-tuning of the visualization, whose main objective consists in making the overview interpretable and not obfuscated by the amount and diversity of information depicted. The assumption behind this mode of operation is the human eye is able to discern and filter patterns in a complex image (Bertin 1967) if guided properly, especially concerning the distinction between various strata and the apprehension of color differences. To this end, the map in Figure 1 is customized with CartoCSS: multiple trial-and-error iterations are performed concerning both data quality and graphical output. The goal resides in finding the most relevant toponyms, which are depicted according to their frequency: only the most prominent ones are labeled, the size of the circles is proportional to the frequency. As expected, Central Europe plays a crucial role in the collection, the existence of numerous data points in Bohemia and Moravia confirms the importance of these regions. The relative importance of points placed in Northern Italy and along the Rhine spark questions about the method and/or their status in the work. They provide a ground for debate and are entry points inasmuch as they are an incentive to explore the digital edition in search for exhaustive evidence. What is more, the toponyms can also be seen as lexical entries to particular themes, especially in the high-frequency range for historical concepts ("Austria-Hungary", "German Empire"), which are placed arbitrarily on the map. In this perspective, mapping the work provides a topical index in a visual form. Since it is exported as an image for the sake of publication, the map can be considered to be static, it cannot be explored dynamically. However, this restriction can be addressed in part by adding a supplementary layer of visual grammar, this time not points but lines.

5  https://github.com/tilemill-project/tilemill
In a further analysis, I visualize co-occurrences of all extracted toponyms, which can be considered to be a subset of GeoCollocations (Bubenhofer 2014), for which the basis has to be a toponym but without restriction on collocates. The two experimental maps below ground on the same data, they result from different settings. In the first case, only toponyms are considered and airborne lines are drawn following their order of appearance. The resulting sequences show which place names appear jointly in the texts.

Maps are mostly projected in Euclidean spaces where two points are connected by a single line. By extension, the word *line* defines a series of connected points on a plane, potentially both linkage and separation. Lines can enforce and divide, in the way that maps are instruments of power, another reason why the proponents of post-representational cartography call for a change of perspective. For example, lines that draw state boundaries are increasingly put into question, especially concerning historical states and texts, as they fix an evolving process and convey a sense of immobility that
fails to describe the past accurately (Smith 2005). The word “network” is to be used with circumspection as Latour (1999) suggests. Although it is ubiquitous in the terminology of the spatial turn, the now predominant interpretation in the sense of the World Wide Web suggests an immediacy which is contrary to the acceptions it had before, so that the concept of “meshwork” is more appropriate (Ingold 2007). Thus, I interpret Figure 2 as a general meshwork which makes it possible to visualize paths depicting chains of thought (Gedankengänge) as well as their intensity (well-trodden or seldom). If they may reveal spatial patterns that would otherwise remain hidden in texts (Bodenhammer et al. 2010), these linkages are also “mappings and tracing imposed on the data” (Wulfman 2014) which are not meant to be interpreted without further filtering.

*Figure 2: Unfiltered map of toponymic co-occurrences*
A rhizome as entry to *Die Fackel*

That is why I refine the map by selecting a subset of the co-occurrences – the maximal distance between two extracted place names is fixed to twenty tokens – and by color-coding qualitative features – the typology of place names (similar to Figure 1) and the axis of time. The most frequent place names are printed out. Surfaces (regions for instance) cannot be represented as such because of historical evolutions and because of the difficulties of linking surfaces without tampering with map readability. Coastlines are depicted in gray to give a sense of orientation, no boundaries are drawn, as they are of a changing nature and may superimpose an artificial reading of the map.

![Figure 3: Refined analysis (size proportional to corpus frequency; yellow: sovereign territories; orange: regions; green: populated places; blue: geographical features; time axis represented by a gradient from light green to dark blue)](image)
In order to interpret Figure 3, I suggest to think about the implications of the concept of rhizome as formalized by Deleuze and Guattari (1980). Its main principles are connection and heterogeneity, with no fixed order; “flat” multiplicity defined by “lines of flight”; “asignifying rupture”; and finally “cartography and decalcomania” (“tracing something that comes ready-made”). The rhizome puts an emphasis on several aspects of post-representational cartography as described above; it “pertains to a map that must be produced, constructed, […] and has multiple entryways and exits” (Deleuze & Guattari 1987 [1980]). For there is no cartographic truth, the map has to be seen as a tool for the multiplication of accesses to reality.

The notion of rhizome has been used in corpus linguistics by Scharloth et al. (2013) to qualify discourses captured by collocation graphs: the authors performed significance tests to extract relevant collocations in the course of time, isolate clusters, and uncover lingering discourse elements. The notion of collocation and of subterranean word complex is a common ground, however the present study is different in its form and its content, because it literally leads to a map and most importantly because its text base is in itself rhizomatic.

This concept is indeed particularly adequate for Kraus, as the Austrian satirist has always been concerned by the multiple aspects of discourse and reality. In addition, his work in Die Fackel evades distant reading processes because of the number of citations used and its ever present and extensive usage of parody. It would be vain to design an authoritative cartography of Die Fackel: following from the principles of heterogeneity and “asignifying rupture” (ibid.), the lines are frequently interrupted. Phenomena in the low-frequency range are filtered out by the human eye, but clusters and interpretation cues may emerge which provide a different access to the work. In this regard, Figure 3 depicts a rhizome connecting heterogeneous information, just as we are all “traversed by lines, geodesics, tropics, and zones marching to different beats and differing in nature” (ibid.).

Conclusion

I have presented a distant reading experiment which consists in connecting toponyms extracted and projected on maps. The latter are meant to be used as additional feature to the existing digital edition. The first map provides a geographical and topical an overview of the collection, while further analyses introduce a sense of dynamism by targeting toponymic cooccurrences: the first example displays unfiltered lines of thought, whereas the second one grounds on a refined analysis and lets the practical image of a rhizome emerge.

This is not an authoritative cartography of Die Fackel but rather an indirect depiction of the viewpoint of Kraus and his contemporaries. Drawing on Kraus’ vitriolic recording of political life, toponyms in Die Fackel tell a story about the ongoing reconfiguration of Europe. As the map conveys the uncanny sensation that the continent is a field on which points east and west are projected, the lines of force entangle European countries and capitals. Their spatial patterns document an inclination for major cultural centers, whereas the chronological dimension captures a

---

6 Software and gazetteer will be made available under open-source licenses, for development files see https://github.com/adbar/geokelone
major shift towards the end of publication: the force field intensifies as its range narrows, showing both the interplay of major European powers of the time and the emergence of transatlantic (westwards) and transeuropean (eastwards) relationships. This evolution can be read as an intensification of tensions and a prefiguration of other schemes, this time of military nature. The notion of rhizome as exposed in this article is not only insightful in this case but also for the scientific process at work: the rhizomatic nature of text visualizations does justice to both philological and technological understandings and thematizes their essence as multiple, disrupted accesses to digital collections.

Indexical maps are entry points both in the sense of an input which enters computational processes and in the hermeneutic sense, since they prompt for contextualizations and interpretations. In fact, digital literary studies are not mere numeric accounts. If on one hand they deal with detecting, counting, and projecting occurrences, on the other hand they have to include and criticize the distanciation provoked by “blind” computer-based thinking (Barbaresi 2012). Because the rhizome is in part a structuralist notion, it may be computed or even computerized. Thus it is only superficially paradoxical that blind numeric analyses can be applied in order to make linguistic phenomena visible. While Die Fackel criticizes mechanical, instrumental language (Hirt 2002), the “well-informed” linguistic instruments can help materializing dots or sequences. However, this is best performed under human supervision. As quantitative and qualitative analysis, both automatically generated outputs and hand-and-eye techniques, are indispensable (Gregory 2011), the human interventions on the maps as well as the technical competence to do so recreate the hermeneutic circle of the philological tradition.

The finality of text visualizations is neither an apparatus nor a notion of an operational nature. It is instead the substrate of interpretable representations which do not follow data but rather interact with them by putting them in perspective. The difference between data wrangling and research in digital humanities resides precisely in the number and diversity of conceptual and technical filters which are repeatedly applied, consciously or sometimes unknowingly. The chosen approach and its inevitable imperfections has to be brought to light, documented and criticized. In this regard, spatial and digital humanities can hopefully be equipped with the adequate instruments and philological background in order to make computed information visible and apprehensible without adding more blind spots.

References


